



Vithanage, CM., Andrieu, C., & Piechocki, RJ. (2007). Novel reduced-state BCJR algorithms. *IEEE Transactions on Communications*, 55(6), 1144 - 1152. <https://doi.org/10.1109/TCOMM.2007.898830>

Peer reviewed version

Link to published version (if available):  
[10.1109/TCOMM.2007.898830](https://doi.org/10.1109/TCOMM.2007.898830)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Novel Reduced-State BCJR Algorithms

Cheran M. Vithanage, *Student Member, IEEE*, Christophe Andrieu, and Robert J. Piechocki, *Member, IEEE*

**Abstract**—BCJR algorithm is an exact and efficient algorithm to compute the marginal posterior distributions of state variables and pairs of consecutive state variables of a trellis structure. Due to its overwhelming complexity, reduced complexity variations, such as the  $M$ -BCJR algorithm, have been developed. In this paper, we propose improvements upon the conventional  $M$ -BCJR algorithm based on modified active state selection criteria. We propose selecting the active states based on estimates of the fixed-lag smoothed distributions of the state variables. We also present Gaussian approximation techniques for the low-complexity estimation of these fixed-lag smoothed distributions. The improved performance over the  $M$ -BCJR algorithm is shown via computer simulations.

**Index Terms**—Decoding, digital communication, fading channels, multiple-input multiple-output (MIMO) systems, nonlinear detection, signal detection, state space methods.

## I. INTRODUCTION

OPTIMAL solution of many problems in digital communications can be formulated as a problem related to the state variables of a discrete-time finite-state Markov process observed in memoryless noise and, hence, as the problem of decoding a trellis structure [1]. An algorithm to compute the marginal posterior distributions of state variables and pairs of consecutive state variables of the trellis with a complexity which is linear with the number of trellis stages (say  $T$ ) is the BCJR algorithm [2], also known as the forward-backward algorithm [3]. Observing that the trellis structure represents an underlying hidden Markov model (HMM) between the state variables and the outputs of the system, one can equivalently apply Pearl's Belief propagation [4] on the corresponding Bayesian network or the sum-product algorithm [5] on the factor graph of the system to compute these marginal distributions with complexities which are linear in  $T$ . Still, the complexity of any such efficient algorithm is also related to the cardinality of the set of realizations of the state variables (say  $\Upsilon$ ) and, hence, is not attractive for practical implementations. For example, in the case of trellis-based equalization of wireless communication systems employed in multiple-input multiple-output (MIMO) channels,  $\Upsilon$  can easily reach millions, which is a huge barrier to reap the capacity potential offered by MIMO channels [6] when

employed in frequency selective channels. Therefore, many reduced complexity variations of the optimal algorithm have been developed, which attempt to reduce complexity possibly with an acceptable reduction in performance. A popular complexity reduction method consists of grouping sets of equalizer states together to build super states such that the modified trellis has state variables taking realizations from a set of lower cardinality. Grouping can be based on truncating the memory of the channel [7], [8] or by set partitioning the symbols input to the channel represented by the trellis [9]. These works are following similar schemes applied for sequence estimation as in [10]–[12]. A trellis splicing scheme is introduced in [13] to remove sections of the trellis in which symbols have been detected with high probability, which can be used in an iterative application of the BCJR algorithm. Franz and Anderson present variations known as  $M$ -BCJR and T-BCJR algorithms in [14], which select a set of states at each trellis stage of the forward recursion as active states and continues computations only from these selected active states. The  $M$ -BCJR algorithm is based on the application of the  $M$  algorithm which has been used for source coding [15] and for Viterbi type sequence decoding [16], with a performance which is highly dependent on the channel energy dispersion [17].

In this paper, we propose enhancements to the  $M$ -BCJR algorithm, resulting in algorithms which are more robust and computationally more efficient for very large trellises. Specifically, we propose to make the active state selection based on an approximation of the fixed-lag smoothed distribution of the state variables. Exact computation of these distributions has a complexity which is also related to  $\Upsilon$  [18], [19]. Hence, we also present Gaussian approximation techniques based on the principle of probabilistic data association (PDA) [20], for the low-complexity estimation of these distributions. The PDA principle has been introduced to communications research in [21] for synchronous multi user detection of CDMA systems and has also been applied for MIMO symbol detection in [22] and soft decision equalization of MIMO systems in [23]. We note here that, in communications literature, it is mostly the aspect of matching the moments of a Gaussian distribution to a more complex distribution which is extracted from the principle of probabilistic data association [21], [23], which originated in target tracking research.

We will present the proposed algorithms in the context of trellis-based equalization of a spatial multiplexing system. The next section will describe a typical structure of such a wireless communication system, and the application of the BCJR and  $M$ -BCJR algorithms will be briefly described in Section III. The proposed algorithms are presented in Section IV with a complexity comparison made in Section V. Finally, computer simulation results and conclusions are given in Sections VI and VII, respectively. Simulation results show the successful application

Paper approved by C. Schlegel, the Editor for Coding Theory and Techniques of the IEEE Communications Society. Manuscript received September 22, 2005; revised August 2, 2006. This work was supported by Toshiba Research Europe, Ltd., Bristol, U.K.

C. M. Vithanage and C. Andrieu are with the Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K. (e-mail: c.vithanage@bristol.ac.uk; c.andrieu@bristol.ac.uk).

R. J. Piechocki is with the Centre for Communications Research, University of Bristol, Bristol BS8 1UB, U.K., and also with the Wireless and Networks Research Laboratory, Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: r.j.piechocki@bristol.ac.uk).

Digital Object Identifier 10.1109/TCOMM.2007.898830

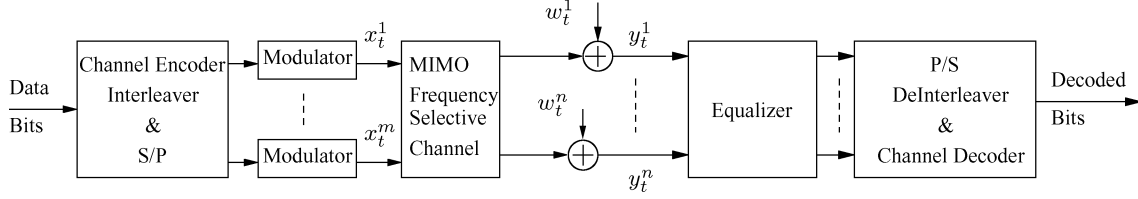


Fig. 1. System model.

of the proposed algorithms on a system with  $\Upsilon = 2^{24}$ , where the comparison is also made with the soft output linear minimum mean squared error equalization technique proposed for MIMO systems in [24]. The ideas and algorithms given here have also been presented in part in [25].

## II. SYSTEM DESCRIPTION

We will consider a spatial multiplexing system with  $m$  transmit antennas and  $n$  receive antennas with a channel having a memory of  $L$  symbol durations as shown in Fig. 1. Considering a discrete time complex baseband analysis, and combining the effects of transmit-receive filters and channel distortions, let the combined channel fading coefficient of the  $d^{\text{th}}$  channel tap from  $j^{\text{th}}$  transmit antenna to the  $i^{\text{th}}$  receive antenna be denoted by  $h^{i,j}(d)$  for  $d \in \{0, \dots, L\}$ ,  $j \in \{1, \dots, m\}$  and  $i \in \{1, \dots, n\}$ . We will consider quasistatic Rayleigh fading channels and perfect channel state information (CSI) at the receiver.

Let  $x_t^j$  and  $y_t^i$  denote the symbol transmitted by the antenna  $j$  and signal received by the antenna  $i$ , respectively, at the  $t^{\text{th}}$  time instant. Let us assume that the symbol  $x_t^j$  is chosen from the alphabet  $B = \{a_1, a_2, \dots, a_N\}$  with cardinality  $N$  and let the frame length be  $T$ . The received signal  $y_t^i$  consists of the convolution of the channel impulse response and a sequence of symbols transmitted up to  $L + 1$  time instants and an additive white Gaussian noise  $w_t^i$

$$y_t^i = \sum_{d=0}^L \left[ \sum_{j=1}^m h^{i,j}(d) x_{t-d}^j \right] + w_t^i.$$

Now, denoting the vector transpose operation by  $(\bullet)^\dagger$ , let us denote the received vector at time  $t$  as  $\mathbf{y}_t = (y_t^1, \dots, y_t^n)^\dagger$  and the symbols transmitted by all antennas at time  $t$  by  $\mathbf{x}_t = (x_t^1, x_t^2, \dots, x_t^m)^\dagger$  which is termed a vector symbol. Note that  $\mathbf{w}_t = (w_t^1, w_t^2, \dots, w_t^n)^\dagger$  is a zero mean circular symmetric complex Gaussian random vector. Let us define  $\mathbf{x}_{t-L:t} = (\mathbf{x}_{t-L}^\dagger, \dots, \mathbf{x}_t^\dagger)^\dagger$  and  $\mathbf{H} = (\mathbf{H}(L) \mathbf{H}(L-1) \dots \mathbf{H}(0))$  with  $\mathbf{H}(d)$  denoting the  $n \times m$  matrix with the  $(i, j)^{\text{th}}$  element being  $h^{i,j}(d)$ . Then, we can express the received signal vector of time  $t$  as

$$\mathbf{y}_t = \mathbf{H} \mathbf{x}_{t-L:t} + \mathbf{w}_t. \quad (1)$$

Optimum equalization to minimize the symbol error rate, detects the transmitted vector symbols based on the maxima of the marginal posterior distributions  $p(\mathbf{x}_t | \mathbf{y}_{1:T})$ , where we are using the notation  $\mathbf{v}_{j:k}$  for  $j \leq k$  to denote the sequence  $\mathbf{v}_j, \mathbf{v}_{j+1}, \dots, \mathbf{v}_k$  for some time indexed set of variables or sets

$\{\mathbf{v}_i\}$ . For the sake of brevity, we will be using the term “distribution” rather loosely to refer to a probability density function or a probability mass function, with the distinction being apparent from the context. Also, we will refer to distributions of the form  $p(\mathbf{v}_t | \mathbf{y}_{1:t+L'})$  with  $L' > 0$  as fixed-lag smoothed distributions, distributions of the form  $p(\mathbf{v}_t | \mathbf{y}_{1:t})$  as filtered distributions and those of the form  $p(\mathbf{v}_t | \mathbf{y}_{1:T})$  as fixed-interval smoothed distributions or simply as posterior distributions.

Considering the subsequent channel decoder, the optimum outputs of the equalizer are the marginal posterior distributions  $p(\mathbf{x}_t | \mathbf{y}_{1:T})$  themselves. To compute these, it is convenient to define a state variable for the equalizer at time instant  $t$  as  $\mathbf{s}_t = (\mathbf{x}_t^\dagger, \dots, \mathbf{x}_{t-L+1}^\dagger)^\dagger$  with a configuration space  $\Psi = B^{mL}$  (note that a trellis is a graphical representation of the realizations of these state variables.). Such a definition ensures that the corresponding finite state machine representation of the equalizer has one of  $\Upsilon = N^{mL}$  possible states at each time instant with an underlying Markov process and with outputs of the system being subject to memoryless noise; and the BCJR algorithm can be applied as follows to compute these marginal posterior distributions. The fact that the channel output at time  $t$  is probabilistically independent of other state variables given the state variables  $\mathbf{s}_{t-1}$  and  $\mathbf{s}_t$  can also be observed from (1). We will reserve the term “state” to denote a particular configuration of a state variable at some time instant.

## III. BCJR AND $M$ -BCJR ALGORITHMS

Let us define

$$\begin{aligned} \alpha(\mathbf{s}_t) &= p(\mathbf{s}_t, \mathbf{y}_{1:t}) \\ \gamma(\mathbf{s}_{t-1}, \mathbf{s}_t) &= p(\mathbf{s}_t | \mathbf{s}_{t-1}) p(\mathbf{y}_t | \mathbf{s}_{t-1}, \mathbf{s}_t) \\ \beta(\mathbf{s}_t) &= p(\mathbf{y}_{t+1:T} | \mathbf{s}_t). \end{aligned}$$

In its operation, the BCJR algorithm involves a forward recursion through the trellis formed by the possibilities of the state variable sequence  $\mathbf{s}_{0:T}$  ( $\mathbf{s}_0$  denotes the initial state variable) in which the recursive computations

$$\alpha(\mathbf{s}_t) = \sum_{\mathbf{s}_{t-1} \in \Psi} \alpha(\mathbf{s}_{t-1}) \gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)$$

are made for  $\mathbf{s}_t \in \Psi, t \in \{1, \dots, T\}$  and a backward recursion in which the computations

$$\beta(\mathbf{s}_t) = \sum_{\mathbf{s}_{t+1} \in \Psi} \beta(\mathbf{s}_{t+1}) \gamma(\mathbf{s}_t, \mathbf{s}_{t+1})$$

are made for  $\mathbf{s}_t \in \Psi, t \in \{1, \dots, T-1\}$  with suitable initializations for  $\alpha(\mathbf{s}_0)$  and  $\beta(\mathbf{s}_T)$ . Due to the flow of  $\alpha(\mathbf{s}_t)$  and  $\beta(\mathbf{s}_t)$  through the trellis diagram, we can call these the forward and backward “messages” passed through the trellis during the

forward and backward recursions, respectively. These messages essentially represent terms which can be shared and factored out in the marginalization of the joint distribution  $p(\mathbf{s}_{0:T}|\mathbf{y}_{1:T})$  to derive the marginals  $p(\mathbf{s}_{t-1}, \mathbf{s}_t|\mathbf{y}_{1:T})$  and  $p(\mathbf{s}_t|\mathbf{y}_{1:T})$  for each  $t \in \{1, \dots, T\}$  [5]. The marginal posterior distributions of state variables and pairs of consecutive state variables (which we can call state transition variables) become available during the backward recursion as

$$\begin{aligned} p(\mathbf{s}_t|\mathbf{y}_{1:T}) &\propto \alpha(\mathbf{s}_t)\beta(\mathbf{s}_t) \\ p(\mathbf{s}_{t-1}, \mathbf{s}_t|\mathbf{y}_{1:T}) &\propto \alpha(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)\beta(\mathbf{s}_t). \end{aligned}$$

Finally, letting  $\zeta(\mathbf{x}_t)$  denote the set of pairs  $(\mathbf{s}_{t-1}, \mathbf{s}_t)$  that lead to the transmission of the vector symbol  $\mathbf{x}_t$ , the posterior distributions of the vector symbols at time instant  $t$  can be computed as

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) \propto \sum_{(\mathbf{s}_{t-1}, \mathbf{s}_t) \in \zeta(\mathbf{x}_t)} \alpha(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)\beta(\mathbf{s}_t).$$

The  $M$ -BCJR algorithm is a reduced complexity variation of this where the forward messages at each time  $t$  are computed for  $M (\ll \Upsilon)$  possible states (which are termed as “active states”) only. The messages of the inactive states are set to zero. For a given set of active states at time  $t-1$  (say  $\Omega_{t-1}$ ), the candidate set for active states at time  $t$  (say  $\Xi_t$ ) is given by the set of states from time  $t$  that can be reached by state transitions from the set  $\Omega_{t-1}$ . The cardinality of  $\Xi_t$ ,  $|\Xi_t| \leq \Upsilon$ , with the inequality being the usual case. At each time instant  $t$ , given the approximate forward messages of time  $t-1$ ,  $\tilde{\alpha}(\mathbf{s}_{t-1})$ ,  $\mathbf{s}_{t-1} \in \Omega_{t-1}$ , the  $M$ -BCJR algorithm computes temporary forward messages for states in  $\Xi_t$  as

$$\hat{\alpha}(\mathbf{s}_t) = \sum_{\mathbf{s}_{t-1} \in \Omega_{t-1}} \tilde{\alpha}(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1}, \mathbf{s}_t) \quad (2)$$

and selects the  $M$  states with the largest messages and makes them the set  $\Omega_t$ . The messages of states in  $\Omega_t$  are kept intact and the other states are considered inactive, i.e.,

$$\begin{cases} \tilde{\alpha}(\mathbf{s}_t) = \hat{\alpha}(\mathbf{s}_t); & \mathbf{s}_t \in \Omega_t \\ \tilde{\alpha}(\mathbf{s}_t) = 0; & \mathbf{s}_t \notin \Omega_t. \end{cases} \quad (3)$$

This recursive selection of the active states results in the selection of a state set sequence  $\Omega_{1:T}$  through the trellis. The backward recursion and the final marginal posterior probabilities are only computed for the states in  $\Omega_{1:T}$ . Specifically, given the backward messages of time  $t+1$  as  $\tilde{\beta}(\mathbf{s}_{t+1})$ ,  $\mathbf{s}_{t+1} \in \Omega_{t+1}$ , the  $M$ -BCJR algorithm computes backward messages and the approximations of the posterior distributions at time  $t$  as

$$\begin{aligned} \tilde{\beta}(\mathbf{s}_t) &= \sum_{\mathbf{s}_{t+1} \in \Omega_{t+1}} \tilde{\beta}(\mathbf{s}_{t+1})\gamma(\mathbf{s}_t, \mathbf{s}_{t+1}); & \mathbf{s}_t \in \Omega_t \\ \tilde{p}(\mathbf{x}_t|\mathbf{y}_{1:T}) &\propto \sum_{(\mathbf{s}_{t-1}, \mathbf{s}_t) \in \xi(\mathbf{x}_t)} \tilde{\alpha}(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)\tilde{\beta}(\mathbf{s}_t). \end{aligned}$$

Here,  $\xi(\mathbf{x}_t)$  denotes the set of pairs  $(\mathbf{s}_{t-1}, \mathbf{s}_t)$  such that  $\mathbf{s}_{t-1} \in \Omega_{t-1}$  and  $\mathbf{s}_t \in \Omega_t$  and leading to the transmission of  $\mathbf{x}_t$ .

Observing that  $\alpha(\mathbf{s}_t) \propto p(\mathbf{s}_t|\mathbf{y}_{1:t})$ , we note that the active state selection of the  $M$ -BCJR algorithm is based on an estimation of the filtered probability distribution of the state variables.

We propose [25] making the active state selection of each time instant based on an estimation of the fixed-lag smoothed probability distribution of the state variable,  $p(\mathbf{s}_t|\mathbf{y}_{1:t+L'})$  with a lag  $L' > 0$ , as described in the next section.

#### IV. PROPOSED ALGORITHMS

Fixed-lag smoothed distributions of the state variables are related to the fixed-lag smoothed distributions of pairs of consecutive state variables as

$$p(\mathbf{s}_t|\mathbf{y}_{1:t+L'}) = \sum_{\mathbf{s}_{t-1} \in B^{mL}} p(\mathbf{s}_{t-1}, \mathbf{s}_t|\mathbf{y}_{1:t+L'}).$$

In order to develop our algorithms, we are going to use three decompositions of  $p(\mathbf{s}_{t-1}, \mathbf{s}_t|\mathbf{y}_{1:t+L'})$ . Using Bayes' formula and noting that given the state variables  $\mathbf{s}_{t-1}$  and  $\mathbf{s}_t$ ,  $\mathbf{y}_{1:t-1}$ ,  $\mathbf{y}_t$  and  $\mathbf{y}_{t+1:t+L'}$  become independent, and taking advantage of the Markov chain structure between the state variables, we observe

$$\begin{aligned} p(\mathbf{s}_{t-1}, \mathbf{s}_t|\mathbf{y}_{1:t+L'}) \\ \propto \alpha(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)p(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_t) \end{aligned} \quad (4)$$

$$\begin{aligned} p(\mathbf{s}_{t-1}, \mathbf{s}_t|\mathbf{y}_{1:t+L'}) \\ \propto \alpha(\mathbf{s}_{t-1})p(\mathbf{y}_t|\mathbf{s}_{t-1}, \mathbf{s}_t)p(\mathbf{x}_t|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'}) \\ \times p(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_{t-1}) \end{aligned} \quad (5)$$

$$\begin{aligned} p(\mathbf{s}_{t-1}, \mathbf{s}_t|\mathbf{y}_{1:t+L'}) \\ \propto \alpha(\mathbf{s}_{t-1})p(\mathbf{x}_t|\mathbf{s}_{t-1}, \mathbf{y}_{t:t+L'})p(\mathbf{y}_{t:t+L'}|\mathbf{s}_{t-1}). \end{aligned} \quad (6)$$

Here, we have also made use of the equivalence in probability of  $p(\mathbf{x}_t|\mathbf{s}_{t-1})$  and  $p(\mathbf{s}_t|\mathbf{s}_{t-1})$  when the state transition from  $\mathbf{s}_{t-1}$  to  $\mathbf{s}_t$  transmits the vector symbol  $\mathbf{x}_t$ . As we shall see below, the decompositions (4)–(6) will present a progressive reduction in implementation complexities especially when the distributions concerned are to be estimated using Gaussian approximation-based methods presented in Section IV-E.

Apart from the FL-MBCJR-1B algorithm, the proposed algorithms differ from the  $M$ -BCJR only in the active state selection phase and have the same backward recursion as the conventional  $M$ -BCJR algorithm.

##### A. FL-MBCJR-1A Algorithm

In this first version of the algorithm, we make use of the decomposition (4). Let us assume that we have computed the forward messages  $\tilde{\alpha}(\mathbf{s}_{t-1})$  for  $\mathbf{s}_{t-1} \in \Omega_{t-1}$  which will also determine the set  $\Xi_t$ . Using (4) and computing the temporary forward messages at time  $t$  as  $\hat{\alpha}(\mathbf{s}_t)$  for  $\mathbf{s}_t \in \Xi_t$  using (2), we can obtain the fixed-lag smoothed distribution of the state variable  $\mathbf{s}_t$  as

$$\tilde{p}(\mathbf{s}_t|\mathbf{y}_{1:t+L'}) \propto \hat{\alpha}(\mathbf{s}_t)\tilde{p}(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_t). \quad (7)$$

Here,  $\tilde{p}(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_t)$  is possibly an approximation of  $p(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_t)$  for each  $\mathbf{s}_t \in \Xi_t$ . One such approximation, particularly well suited for our equalization example as demonstrated by simulations, is presented in Section IV-E1.

Thus, we find the modified forward recursion at time instant  $t$  (for  $t \in \{1, \dots, T\}$ ) as follows.

- Compute  $\gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)$  and  $\hat{\alpha}(\mathbf{s}_t)$  for  $\mathbf{s}_{t-1} \in \Omega_{t-1}$  and  $\mathbf{s}_t \in \Xi_t$  using (2).
- Compute  $\tilde{p}(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_t)$  for  $\mathbf{s}_t \in \Xi_t$ .

- Compute  $\tilde{p}(s_t|y_{1:t+L'})$  using (7) for  $s_t \in \Xi_t$ . (The normalization can be discarded.)
- Select the  $M$  states of  $\Xi_t$  which have the largest estimated fixed-lag smoothed probability, and make them the set  $\Omega_t$ .
- Obtain the forward messages at time  $t$  using (3).

It is possible to make use of the availability of  $\hat{\alpha}(s_t)$  for  $s_t \in \Xi_t$  (instead of just  $\tilde{\alpha}(s_t)$  for  $s_t \in \Omega_t$ ) in estimating the final posterior marginal distributions of the symbols, at the cost of increasing the computational complexity. We also note here that both the computations  $\hat{\alpha}(s_t)$  and  $\tilde{p}(y_{t+1:t+L'}|s_t)$  need to be performed for every  $s_t \in \Xi_t$ .

### B. FL-MBCJR-1B Algorithm

Given that we can obtain  $p(s_{t-1}, s_t|y_{1:t+L'})$  by

$$\tilde{p}(s_{t-1}, s_t|y_{1:t+L'}) \propto \tilde{\alpha}(s_{t-1})\gamma(s_{t-1}, s_t)\tilde{p}(y_{t+1:t+L'}|s_t) \quad (8)$$

we can discard the backward recursion and develop a purely fixed-lag smoothing algorithm which has a complexity independent of  $\Upsilon$ . This variation is more suitable for reduced complexity fixed-delay trellis decoding (with a delay of  $L'$ ). In this algorithm, the fixed-lag smoothed distributions of the transmitted symbols are computed as

$$\tilde{p}(x_t|y_{1:t+L'}) \propto \sum_{(s_{t-1}, s_t) \in \xi(x_t)} \tilde{p}(s_{t-1}, s_t|y_{1:t+L'}). \quad (9)$$

The FL-MBCJR-1B algorithm which consists of a forward recursion only, performs the steps given below for each time instant  $t$ .

- Compute  $\tilde{p}(y_{t+1:t+L'}|s_t)$  and  $\gamma(s_{t-1}, s_t)$  for  $s_{t-1} \in \Omega_{t-1}$  and  $s_t \in \Xi_t$ .
- Compute  $\tilde{p}(s_{t-1}, s_t|y_{1:t+L'})$  using (8) for  $s_{t-1} \in \Omega_{t-1}$  and  $s_t \in \Xi_t$ .
- Compute the soft output  $\tilde{p}(x_t|y_{1:t+L'})$  for  $x_t \in B^m$  using (9).
- Compute  $\tilde{p}(s_t|y_{1:t+L'})$  for  $s_t \in \Xi_t$  by summing (8) over  $s_{t-1} \in \Omega_{t-1}$ .
- Select the  $M$  states of  $\Xi_t$  which have the largest estimated fixed-lag smoothed probability, and make them the set  $\Omega_t$ .
- Compute the temporary forward messages  $\tilde{\alpha}(s_t)$  for states in  $\Omega_t$  only.

The absence of the backward recursion results in the reduction in complexity compared to the FL-MBCJR-1A algorithm.

### C. FL-MBCJR-2A Algorithm

Using the decomposition (5) we can obtain the fixed-lag smoothed distribution of the state variable  $s_t$  by

$$\tilde{p}(s_t|y_{1:t+L'}) \propto \sum_{s_{t-1} \in \Omega_{t-1}} \tilde{\alpha}(s_{t-1})p(y_t|s_{t-1}, s_t) \cdot \tilde{p}(x_t|s_{t-1}, y_{t+1:t+L'})\tilde{p}(y_{t+1:t+L'}|s_{t-1}). \quad (10)$$

Here,  $\tilde{p}(y_{t+1:t+L'}|s_{t-1})$  is possibly an approximation of  $p(y_{t+1:t+L'}|s_{t-1})$  obtained as in Section IV-E1 (now the conditioning is on some state  $s_{t-1} \in \Omega_{t-1}$ ). Likewise,  $\tilde{p}(x_t|s_{t-1}, y_{t+1:t+L'})$  is possibly an approximation of  $p(x_t|s_{t-1}, y_{t+1:t+L'})$ . Section IV-E2 presents how this proba-

bility can be computed exactly and an approximation method which performed well in the simulations.

The modified forward recursion at time instant  $t$  is as follows.

- Compute  $\tilde{p}(y_{t+1:t+L'}|s_{t-1})$  and  $p(y_t|s_{t-1}, s_t)$  for  $s_{t-1} \in \Omega_{t-1}$  and  $s_t \in \Xi_t$ .
- Compute  $\tilde{p}(x_t|s_{t-1}, y_{t+1:t+L'})$  for  $s_{t-1} \in \Omega_{t-1}$ .
- Compute  $\tilde{p}(s_t|y_{1:t+L'})$  using (10) for  $s_t \in \Xi_t$  (the normalization can be discarded).
- Select the  $M$  states of  $\Xi_t$  which have the largest estimated fixed-lag smoothed probability, and make them the set  $\Omega_t$ .
- Compute  $\gamma(s_{t-1}, s_t)$  for  $s_{t-1} \in \Omega_{t-1}$  and  $s_t \in \Omega_t$  only.
- Compute the new forward messages  $\tilde{\alpha}(s_t)$  for states in  $\Omega_t$  only.

In (10), we can note that apart from  $p(y_t|s_{t-1}, s_t)$ , all the probability distributions are conditioned on an active state at time  $t-1$  instead of on a candidate for an active state at time  $t$ . The fact that the number of states in  $\Omega_{t-1}$  (which is  $M$ ) is much less than the number of states in  $\Xi_{t-1}$  (which can be up to  $MN^m$ ) results in a reduction in complexity compared to the earlier algorithms, especially when the number of transmit antennas increases.

### D. FL-MBCJR-2B Algorithm

We can use (6) to obtain the fixed-lag smoothed distribution of the state variable  $s_t$  by

$$\tilde{p}(s_t|y_{1:t+L'}) \propto \sum_{s_{t-1} \in \Omega_{t-1}} \tilde{\alpha}(s_{t-1})\tilde{p}(x_t|s_{t-1}, y_{t:t+L'}) \cdot \tilde{p}(y_{t:t+L'}|s_{t-1}). \quad (11)$$

Again,  $\tilde{p}(y_{t:t+L'}|s_{t-1})$  is possibly an approximation of  $p(y_{t:t+L'}|s_{t-1})$  obtained as in Section IV-E1 (now the conditioning is on some state  $s_{t-1} \in \Omega_{t-1}$  and the received values in time instants  $\{t, \dots, t+L'\}$  need to be considered). In this algorithm, we will always make the approximation of assuming the symbols transmitted by different antennas at each time instant to be conditionally independent given the observations through the channel [23]

$$\tilde{p}(x_t|s_{t-1}, y_{t:t+L'}) \approx \prod_{j=1}^m \tilde{p}(x_t^j|s_{t-1}, y_{t:t+L'}). \quad (12)$$

The estimation of  $\tilde{p}(x_t^j|s_{t-1}, y_{t:t+L'})$  is made as in Section IV-E2 (considering the received values in time instants  $\{t, \dots, t+L'\}$ ).

Furthermore, for each  $s_{t-1} \in \Omega_{t-1}$ , we will make use of the approximation  $\tilde{p}(x_t|s_{t-1}, y_{t:t+L'})$  and select only  $S (\leq N^m)$  states from time  $t$  to be included in  $\Xi_t$  [note that the factorization of  $\tilde{p}(x_t|s_{t-1}, y_{t:t+L'})$  in (12) enables this step to have a linear complexity in  $m$ ]. In other words, we will also select state transitions from the set of active states of time  $t-1$ .

Therefore, the modified forward recursion at time instant  $t$  is as follows.

- Compute  $\tilde{p}(y_{t:t+L'}|s_{t-1})$  for  $s_{t-1} \in \Omega_{t-1}$ .
- Compute  $\tilde{p}(x_t^j|s_{t-1}, y_{t:t+L'})$  for  $j \in \{1, \dots, m\}$  and  $s_{t-1} \in \Omega_{t-1}$ .
- Compute the  $S \leq N^m$  largest components of  $\tilde{p}(x_t|s_{t-1}, y_{t:t+L'})$  using (12) for each  $s_{t-1} \in \Omega_{t-1}$  and include the corresponding  $s_t$  in  $\Xi_t$ .

- Compute  $\tilde{p}(\mathbf{s}_t | \mathbf{y}_{1:t+L'})$  using (11) for  $\mathbf{s}_t \in \Xi_t$  (the normalization can be neglected).
- Select the  $M$  states of  $\Xi_t$  which have the largest estimated fixed-lag smoothed probability, and make them the set  $\Omega_t$ .
- Compute  $\gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)$  for  $\mathbf{s}_{t-1} \in \Omega_{t-1}$  and  $\mathbf{s}_t \in \Omega_t$  only.
- Compute the new forward messages  $\tilde{\alpha}(\mathbf{s}_t)$  for states in  $\Omega_t$  only.

In (11), we can note that all the probability distributions are conditioned on an active state at time  $t - 1$  instead on a candidate for an active state at time  $t$ . Therefore, this version presents the lowest computational complexity which is in fact linear in the number of transmit antennas with the additional help of the approximation (12). It is also straightforward to implement a fixed-delay decoding variation for this algorithm, following the procedure of FL-MBCJR-1B.

### E. Exact and Moment-Matching-Based Approximate Computation of Intermediate Likelihoods and Probability Distributions

The received values  $\mathbf{y}_{t+1:t+L'}$  and the transmitted modulation symbols that have energy components returned within time  $[t + 1, t + L']$  can be related as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \quad (13)$$

where  $\mathbf{x} = (\mathbf{x}_{t-L+1}^\dagger, \dots, \mathbf{x}_{t+L'}^\dagger)^\dagger$ ,  $\mathbf{y} = (\mathbf{y}_{t+1}^\dagger, \dots, \mathbf{y}_{t+L'}^\dagger)^\dagger$  and  $\mathbf{w} = (\mathbf{w}_{t+1}^\dagger, \dots, \mathbf{w}_{t+L'}^\dagger)^\dagger$ . The  $nL' \times m(L + L')$  matrix  $\mathbf{A}$  is given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{H}(L) & \mathbf{H}(L-1) & \dots & \mathbf{H}(0) & \dots & 0 \\ 0 & \mathbf{H}(L) & \dots & \mathbf{H}(1) & \mathbf{H}(0) & 0 \\ 0 & 0 & & \ddots & \ddots & 0 \\ 0 & 0 & & \mathbf{H}(L) & \dots & \mathbf{H}(0) \end{bmatrix}$$

where  $\mathbf{H}(d)$  denotes the  $n \times m$  matrix with the  $(i, j)^{th}$  element being  $h_{i,j}^d(d)$ . In the matrix multiplication  $\mathbf{A}\mathbf{x}$  of (13), there is a column of  $\mathbf{A}$  involved with the element  $x_k^l$ . Let us denote that column by  $\mathbf{h}_k^l$ . Now (13) can be rewritten as

$$\mathbf{y} = \sum_{\substack{k \in \{t-L+1, \dots, t\} \\ l \in \{1, 2, \dots, m\}}} \mathbf{h}_k^l x_k^l + \sum_{\substack{k \in \{t+1, \dots, t+L'\} \\ l \in \{1, 2, \dots, m\}}} \mathbf{h}_k^l x_k^l + \mathbf{w}.$$

1) *Computation of  $\tilde{p}(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t)$* : For a given state  $\mathbf{s}_t \in \Xi_t$ ,  $x_k^l$  is known only for  $k = t - L + 1, \dots, t$  and  $l = 1, 2, \dots, m$ . Let  $\mathbf{s}_{-t}$  denote the set of transmitted symbols  $x_k^l$  for  $k = t + 1, \dots, t + L'$  and  $l = 1, 2, \dots, m$ . Now making use of the independence of the transmitted symbols, we can see that  $p(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t)$  can be exactly computed as

$$p(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t) = \sum_{\mathbf{s}_{-t}} p(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t, \mathbf{s}_{-t}) p(\mathbf{s}_{-t}) \quad (14)$$

where we can make use of the fact that  $p(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t, \mathbf{s}_{-t})$  is simply a multivariate complex Gaussian distribution. The prior distribution  $p(\mathbf{s}_{-t})$  can be considered to be uniform in the absence of any information on them. This derivation involves an enumeration over the possibilities of  $\mathbf{s}_{-t}$  and, hence, has a complexity which is exponential in  $m$  and  $L'$ . As this complexity can

be unacceptable for large  $L'$ , we present the following example of an approximation procedure.

As seen from (14), the distribution  $p(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t)$  is actually a mixture of Gaussians. The evaluation of  $p(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t)$  for the particular received values can be simplified using a Gaussian approximation. We will use the PDA concept of approximating a complex probability distribution such as a mixture of Gaussians by a single moment matched Gaussian distribution, to approximate the Gaussian mixture in our case. We note here that a multivariate complex Gaussian distribution is completely specified by its mean vector, covariance matrix and the pseudo-covariance matrix [26]. Assuming that the transmitted antenna symbols are independent (which is justified due to the interleaver before the spatial multiplexer in this case) the mean, covariance matrix and the pseudo-covariance matrix of the distribution of random variable  $\mathbf{y}$  given  $\mathbf{s}_t$ ,  $\bar{\boldsymbol{\mu}}$ ,  $\bar{\boldsymbol{\Sigma}}$ , and  $\bar{\mathbf{J}}$  are given by

$$\begin{aligned} \bar{\boldsymbol{\mu}} &= \sum_{\substack{k \in \{t-L+1, \dots, t\} \\ l \in \{1, 2, \dots, m\}}} \mathbf{h}_k^l x_k^l + \sum_{\substack{k \in \{t+1, \dots, t+L'\} \\ l \in \{1, 2, \dots, m\}}} \mathbf{h}_k^l E(x_k^l) \\ \bar{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma} + \sum_{\substack{k \in \{t+1, \dots, t+L'\} \\ l \in \{1, 2, \dots, m\}}} \mathbf{h}_k^l (\mathbf{h}_k^l)^\dagger E(|x_k^l - E(x_k^l)|^2) \\ \bar{\mathbf{J}} &= \sum_{\substack{k \in \{t+1, \dots, t+L'\} \\ l \in \{1, 2, \dots, m\}}} \mathbf{h}_k^l (\mathbf{h}_k^l)^\dagger E((x_k^l - E(x_k^l))^2). \end{aligned}$$

Here,  $(\bullet)^\dagger$  denotes the conjugate transpose operation,  $\boldsymbol{\Sigma}$  is the covariance matrix of  $\mathbf{w}$  given by  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{nL'}$ , with  $\sigma^2$  being the noise variance on each receive antenna and  $\mathbf{I}_p$  denoting the  $p \times p$  identity matrix. The expectations  $E(\bullet)$  are taken with respect to any available prior information on the modulated symbol transmitted by antenna  $l$  at time instant  $k$ , which can be nonuniform such as when used in a turbo equalization scheme or when this algorithm is iterated itself to obtain better defined maxima in the fixed-lag distributions.

Now, we can moment match a single Gaussian distribution to  $p(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t)$ , with matched parameters  $\bar{\boldsymbol{\mu}}$ ,  $\bar{\boldsymbol{\Sigma}}$  and  $\bar{\mathbf{J}}$ , and derive an approximation to the probability  $p(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t)$  as

$$\tilde{p}(\mathbf{y} | \mathbf{s}_t) = \frac{1}{\pi^{nL'} \sqrt{\det(\bar{\mathbf{K}})}} \cdot \exp \left\{ - \begin{pmatrix} \text{Re}(\mathbf{y} - \bar{\boldsymbol{\mu}}) \\ \text{Im}(\mathbf{y} - \bar{\boldsymbol{\mu}}) \end{pmatrix}^\dagger \bar{\mathbf{K}}^{-1} \begin{pmatrix} \text{Re}(\mathbf{y} - \bar{\boldsymbol{\mu}}) \\ \text{Im}(\mathbf{y} - \bar{\boldsymbol{\mu}}) \end{pmatrix} \right\}$$

with

$$\bar{\mathbf{K}} = \begin{pmatrix} \text{Re}(\bar{\mathbf{J}} + \bar{\boldsymbol{\Sigma}}) & \text{Im}(\bar{\mathbf{J}} - \bar{\boldsymbol{\Sigma}}) \\ \text{Im}(\bar{\mathbf{J}} + \bar{\boldsymbol{\Sigma}}) & \text{Re}(\bar{\boldsymbol{\Sigma}} - \bar{\mathbf{J}}) \end{pmatrix}.$$

Here,  $(\bullet)^{-1}$  denotes matrix inversion operation and  $\det(\bar{\mathbf{K}})$  is the determinant of the matrix  $\bar{\mathbf{K}}$ .

We can also note at this point that the inverse and determinant of matrix  $\bar{\mathbf{K}}$  need to be computed only once per frame transmission when there is no prior information on the transmitted symbols. This is the case when the algorithm is used in a serially concatenated equalizer decoder system without any iterations within the equalizer. Otherwise, there is scope for the use

TABLE I  
ORDER OF COMPUTATIONS PER TIME INSTANT

Algorithm	ADD/SUB/COMP	MUL/DIV	Exp
FL-MBCJR-1A	$MN^m \cdot \max\{Lmn, (nL')^2, \log_2(MN^m)\}$	$MN^m \cdot \max\{(L+1)mn, (nL')^2\}$	$MN^m$
FL-MBCJR-2B	$M \cdot \max\{LmnM, n^2(L'+1)^2Nm, S\log_2(MS)\}$	$M \cdot \max\{(L+1)mnM, n^2(L'+1)^2Nm, (m+1)S\}$	$MN^m$
M-BCJR	$MN^m \cdot \max\{Lmn, \log_2(MN^m)\}$	$(L+1)mnMN^m$	$MN^m$

of the matrix inversion lemma to reduce the complexity of computing the matrix inverses at each time instant. If the transmitted symbols belong to a rotationally invariant constellation such as QPSK, 8PSK, or 16QAM in addition to having no prior information, then the pseudo-covariance matrix vanishes leading to the simpler and familiar expression

$$\tilde{p}(\mathbf{y}|\mathbf{s}_t) = \frac{1}{\pi^{nL'} \det(\tilde{\Sigma})} \exp\left(-(\mathbf{y} - \tilde{\mu})^\dagger (\tilde{\Sigma})^{-1} (\mathbf{y} - \tilde{\mu})\right).$$

Otherwise, the approximation of neglecting the pseudo-covariance matrix can be made to reduce the computational complexity.

2) *Computation of  $\tilde{p}(\mathbf{x}_t|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'})$* : For a given state  $\mathbf{s}_{t-1} \in \Omega_{t-1}$ ,  $x_k^l$  is fixed only for  $k = t-L, \dots, t-1$  and  $l = 1, 2, \dots, m$ . Now, making use of the independence of the transmitted symbols we can see that  $p(\mathbf{x}_t|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'})$  can be exactly computed as

$$\tilde{p}(\mathbf{x}_t|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'}) \propto \left[ \sum_{\mathbf{s}_{-t}} p(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_{t-1}, \mathbf{x}_t, \mathbf{s}_{-t}) p(\mathbf{s}_{-t}) \right] p(\mathbf{x}_t).$$

This exact computation can be seen to have a complexity which is exponential in both  $m$  and  $L'$ . One method of approximating these distributions at a much lower complexity is given below.

We will first make the assumption of considering the symbols transmitted by the separate antennas remain independent even after the observations through the channel [23], which will lead to

$$p(\mathbf{x}_t|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'}) \approx \prod_{j=1}^m p\left(x_t^j|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'}\right).$$

Thus, our problem reduces to estimating  $p(x_t^j|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'})$  for  $j \in \{1, \dots, m\}$ . Note that

$$\begin{aligned} p\left(x_t^j|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'}\right) \\ = p\left(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_{t-1}, x_t^j\right) \frac{p\left(x_t^j|\mathbf{s}_{t-1}\right)}{p\left(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_{t-1}\right)}. \end{aligned}$$

Assuming prior independence of the transmitted symbols and uniform prior information, we have

$$p\left(x_t^j|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'}\right) \propto p\left(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_{t-1}, x_t^j\right). \quad (15)$$

Now, let us consider the distribution  $p(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_{t-1}, x_t^j)$ , which is a mixture of Gaussians. In this method of approximation, we will moment match a single Gaussian distribution to this mixture.

We note again that we are assuming perfect CSI at the receiver. For any  $j \in \{1, \dots, m\}$ , we can rewrite (13) as

$$\mathbf{y} = \mathbf{h}_t^j x_t^j + \sum_{\substack{k \in \{t-L+1, \dots, t-1\} \\ l \in \{1, \dots, m\}}} \mathbf{h}_k^l x_k^l + \sum_{\substack{k \in \{t, \dots, t+L'\} \\ l \in \{1, \dots, m\} \\ (k,l) \neq (t,j)}} \mathbf{h}_k^l x_k^l + \mathbf{w}. \quad (16)$$

Let us define

$$\begin{aligned} \hat{\mu} &= \sum_{\substack{k \in \{t-L+1, \dots, t-1\} \\ l \in \{1, \dots, m\}}} \mathbf{h}_k^l x_k^l + \sum_{\substack{k \in \{t, \dots, t+L'\} \\ l \in \{1, \dots, m\} \\ (k,l) \neq (t,j)}} \mathbf{h}_k^l E(x_k^l) \\ \hat{\Sigma} &= \Sigma + \sum_{\substack{k \in \{t, \dots, t+L'\} \\ l \in \{1, \dots, m\} \\ (k,l) \neq (t,j)}} \mathbf{h}_k^l (\mathbf{h}_k^l)^\dagger E\left(|x_k^l - E(x_k^l)|^2\right) \end{aligned}$$

and

$$\hat{\mathbf{J}} = \sum_{\substack{k \in \{t, \dots, t+L'\} \\ l \in \{1, \dots, m\} \\ (k,l) \neq (t,j)}} \mathbf{h}_k^l (\mathbf{h}_k^l)^\dagger E\left((x_k^l - E(x_k^l))^2\right).$$

As in the earlier section, the matched single Gaussian distribution is given by

$$\begin{aligned} \tilde{p}\left(\mathbf{y}_{t+1:t+L'}|\mathbf{s}_{t-1}, x_t^j\right) \\ \propto \exp\left\{-\left(\begin{matrix} \text{Re}(\mathbf{y} - \tilde{\mu}) \\ \text{Im}(\mathbf{y} - \tilde{\mu}) \end{matrix}\right)^\dagger \hat{\mathbf{K}}^{-1} \begin{pmatrix} \text{Re}(\mathbf{y} - \tilde{\mu}) \\ \text{Im}(\mathbf{y} - \tilde{\mu}) \end{pmatrix}\right\} \end{aligned}$$

where  $\tilde{\mu} = (\mathbf{h}_t^j x_t^j - \hat{\mu})$  and

$$\hat{\mathbf{K}} = \begin{pmatrix} \text{Re}(\hat{\mathbf{J}} + \hat{\Sigma}) & \text{Im}(\hat{\mathbf{J}} - \hat{\Sigma}) \\ \text{Im}(\hat{\mathbf{J}} + \hat{\Sigma}) & \text{Re}(\hat{\Sigma} - \hat{\mathbf{J}}) \end{pmatrix}$$

Thus, the approximation to  $p(x_t^j|\mathbf{s}_{t-1}, \mathbf{y}_{t+1:t+L'})$  becomes available due to (15).

## V. COMPLEXITY COMPARISON

The major orders of complexities of the FL-MBCJR-1A, FL-MBCJR-2B, and M-BCJR algorithms are given in Table I. The order of computations of addition/subtraction/comparison (ADD/SUB/COMP) operations, multiplication/division (MUL/DIV) operations and exponential function evaluations are shown. The algorithms are assumed to be implemented non iteratively with no prior information on the symbols. Also the fixed-lag smoothed distributions are assumed to be estimated using the Gaussian approximations presented earlier. In computing the complexities, sorting a vector of length  $R$  was assumed to have a worst case complexity of  $\mathcal{O}(R \log_2 R)$  comparison operations (which is achieved, for example, by using the heapsort algorithm). The exponential function evaluations arise during the computation of the probability of a

realization of a random variable with a Gaussian distribution such as  $p(\mathbf{y}_t | \mathbf{s}_{t-1}, \mathbf{s}_t)$  or  $\gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)$  or  $\tilde{p}(\mathbf{y}_{t+1:t+L'} | \mathbf{s}_t)$  with the moment matched parameters. These exponential function evaluations can be avoided by operating the algorithms in their logarithms, which will also improve numerical stability in practical implementations. We note that the operation in the log domain will require taking logarithms of a sum of exponential functions, for example, when trying to compute (3). The complexity of this step can be reduced with the use of the Jacobian logarithm and the lookup-table-based corrections [27].

Even though the complexities of the FL-MBCJR-1B and FL-MBCJR-2A algorithms are progressively lower compared to the FL-MBCJR-1A under these assumptions, they do not enforce a major reduction in the order of complexity. As can be seen from the Table I, the order of complexity of the FL-MBCJR-2B algorithm is significantly lower, and in fact even lower than that of the  $M$ -BCJR algorithm for the same number of active states for large  $N$  and  $m$ .

## VI. SIMULATION RESULTS

In the computer simulations, the transmission frames consisted of 144 data bits which were channel coded using a rate half turbo code using two constituent  $(5, 7)_8$  convolutional codes. The decoders performed 4 iterations of turbo decoding. The proposed algorithms as well as the BCJR and  $M$ -BCJR algorithms were performed in their logarithms for better numerical stability. Simulations were carried out on quasistatic Rayleigh fading channels including 5-tap channels with the energy distribution of the 5 taps as

$$\text{Channel 1 : } [0.8 \quad 0.12 \quad 0.06 \quad 0.015 \quad 0.005]$$

$$\text{Channel 2 : } [0.0515 \quad 0.2116 \quad 0.4738 \quad 0.2116 \quad 0.0515].$$

For very large state spaces, the proposed algorithms were also compared against the soft output linear minimum mean squared error (MMSE) equalization method for spatial multiplexing systems presented in [24], which is an extension of the scheme of [28] for spatial multiplexing systems. These windowed MMSE schemes were simulated with a forward window of  $2(L+1)$  symbols and a backward window of  $(L+1)$  symbols. The proposed algorithms were run with a fixed-lag of  $3L$  unless otherwise stated.

Fig. 2 shows the simulation in an  $m = n = 2$  system with BPSK transmission into the 5-tap channels Channel 1 and Channel 2. The bit error rate of quantized decisions after the equalizer is plotted. Even though  $\Upsilon = 256$  for this system, it can be seen that both  $M$ -BCJR and FL-MBCJR-1A schemes using only four active states have near optimum performance in Channel 1 where most of the energy of the transmitted symbols is returned in the first tap, in terms of this hard output of the equalizer. In Channel 2, where a significant part of the symbol energy is returned in the middle taps, we can see the conventional  $M$ -BCJR algorithm failing in performance and that the proposed scheme still remains near optimal. Thus, we can clearly observe the benefit of the proposed consideration of the received signals from time  $t+1$  to  $t+L'$  in making the active state selection of time  $t$ .

The effect of the fixed-lag,  $L'$  is observed in Fig. 3. The simulated system has  $m = n = 2$  with 8PSK transmission into

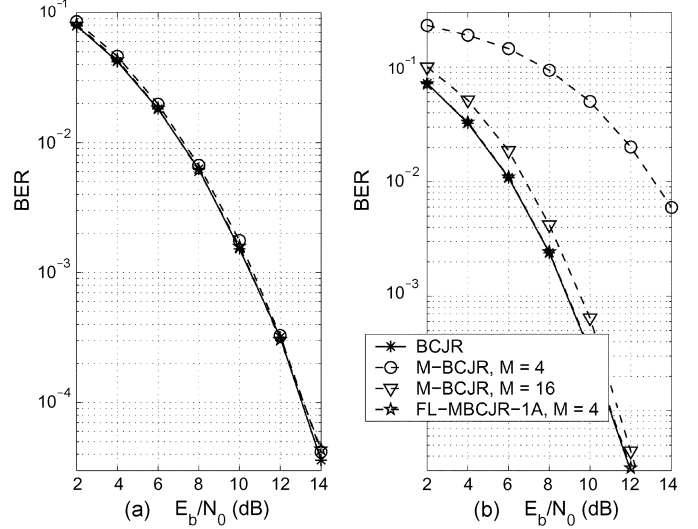


Fig. 2. Uncoded performance of BCJR,  $M$ -BCJR, and FL-MBCJR-1A schemes with  $2 \times 2$  BPSK transmission into (a) Channel 1 and (b) Channel 2.  $\Upsilon = 256$ .

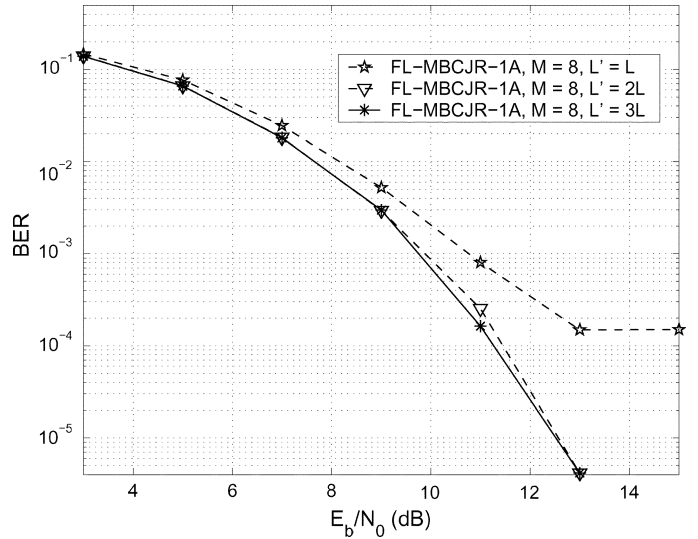


Fig. 3. Coded performance of FL-MBCJR-1A scheme with  $2 \times 2$  8PSK transmission into Channel 2 and different fixed-lag,  $L'$  at the equalizer.  $\Upsilon = 2^{24}$ .

Channel 2. Hence, the number of states of an optimal equalizer is about 16 million. Fixed lags of  $L$ ,  $2L$  and  $3L$  are simulated. It can be seen that the relative improvement in error rate performance with the fixed-lag  $L'$  decreases with  $L'$  and can also be understood in terms of the exponential forgetting in fixed-lag smoothers [29] for hidden Markov models.

Fig. 4 shows the effect of the number of active state transitions of the FL-MBCJR-2B scheme. For this  $m = n = 2$  system employed in Channel 2, transmitting 8PSK modulated symbols on each antenna, the maximum value of  $S$  is  $N^m = 64$ . The computational savings offered by the selection of active state transitions is seen by the almost identical performance for  $S$  values of 4 and 64.

Fig. 5 makes a comparison of the proposed schemes with  $M$ -BCJR and the soft output MMSE scheme in a  $2 \times 2$  system



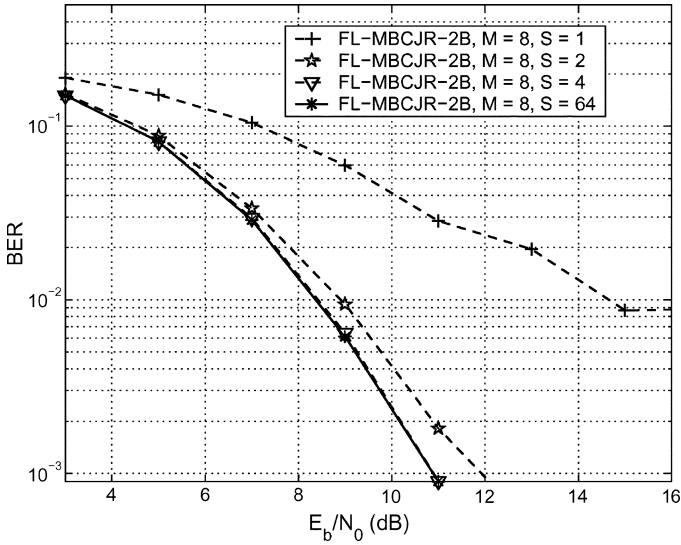


Fig. 4. Coded performance of FL-MBCJR-2B scheme with  $2 \times 2$  8PSK transmission into Channel 2 and different number of active state transitions,  $S$  at the equalizer.  $\Upsilon = 2^{24}$ .

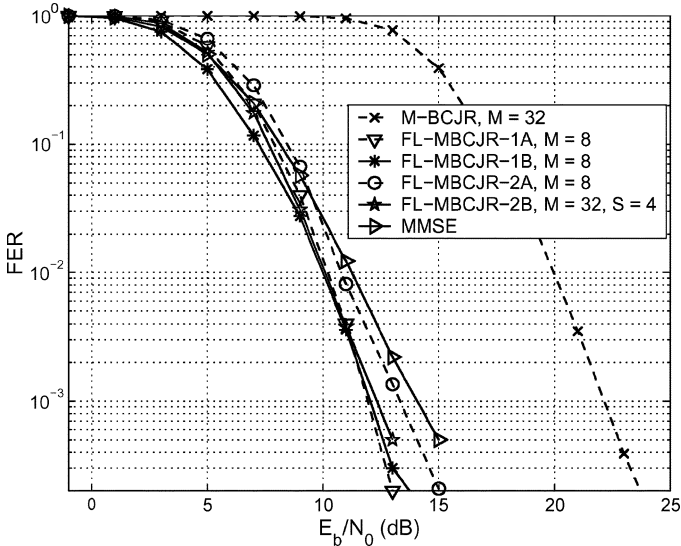


Fig. 5. Coded performance of proposed schemes with  $2 \times 2$  8PSK transmission into Channel 2.  $\Upsilon = 2^{24}$ .

with 8PSK transmission. The proposed schemes with 8–32 active states have a better error rate performance than the conventional  $M$ -BCJR algorithm and the MMSE scheme. Of particular interest is the comparison between the FL-MBCJR-1A and FL-MBCJR-1B schemes, which is further highlighted in Fig. 6. FL-MBCJR-1A is an approximation to a fixed-interval decoding scheme while FL-MBCJR-1B is an approximation to a fixed-delay decoding scheme. It can be seen that at low bit to noise energy ratios, the second scheme outperforms the first. With the increase of transmitted energy, the effect of smoothing takes over, and the FL-MBCJR-1A algorithm outperforms the FL-MBCJR-1B algorithm.

The effect of channel memory  $L$  on the number of active states required for error rate performance compared to the MMSE scheme is investigated in Fig. 7. With  $L = 3$  and  $L = 10$ , the optimal equalizers have 64 and over 134 million active states, respectively. Still, the FL-MBCJR-2B scheme

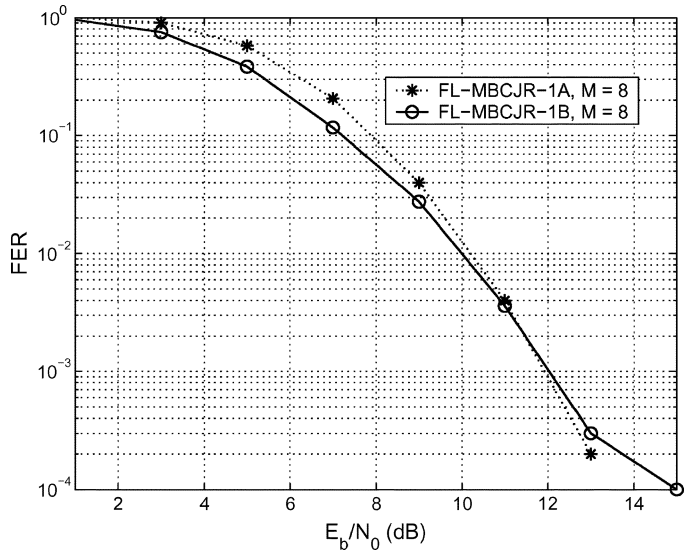


Fig. 6. Coded performance of FL-MBCJR-1A and FL-MBCJR-1B schemes with  $2 \times 2$  8PSK transmission into Channel 2.  $\Upsilon = 2^{24}$ .

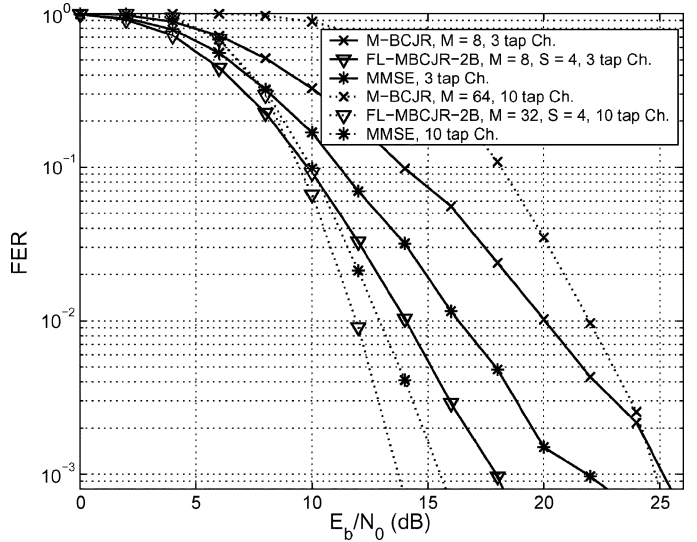


Fig. 7. Coded performance of FL-MBCJR-2B scheme with  $1 \times 1$  8PSK transmission into uniform delay profile channels with memory 3 and 10 having  $\Upsilon = 64$  and  $\Upsilon = 2^{27}$ , respectively.

with  $S = 4$  has a better error rate performance than the MMSE scheme by using only 8 and 32 active states, respectively. The performance improvement over the  $M$ -BCJR algorithm can be seen to be widening with the enlargement of the state space.

## VII. CONCLUSION

We have proposed new trellis decoding algorithms by changing the active state selection criteria of the  $M$ -BCJR algorithm from approximations of filtered distributions of state variables to approximations of fixed-lag smoothed distributions of state variables. Several methods based on Gaussian approximation techniques are given for the estimation of these distributions with low complexity. The performance improvement over the  $M$ -BCJR algorithm is shown via computer simulations, where the new schemes remain robust to changes in the channel multipath profile and the successful application of the proposed algorithms is given for a system with  $\Upsilon = 2^{24}$ .

Of the proposed algorithms, the FL-MBCJR-2B scheme presents the largest computational savings for multiple antenna transmissions. The FL-MBCJR-1B algorithm presents a fixed-delay decoding scheme which can be useful for online decoding of trellises. Noting that the complexities of these algorithms are quadratic in the fixed-lag  $L'$ , one could implement the channel shortening prefiltering method of [17] prior to trellis-based equalization using the proposed algorithms.

The ideas presented here can easily be applied to the T-BCJR algorithm presented in [14], which will lead to larger computational savings at high signal to noise energy ratios.

#### ACKNOWLEDGMENT

The authors would like to thank M. S. Yee, J. P. Coon, and the anonymous reviewers for their valuable comments.

#### REFERENCES

- [1] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.
- [2] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 284–287, Mar. 1974.
- [3] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.
- [5] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [6] E. Telatar, "Capacity of the multi antenna Gaussian channel," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [7] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-state BCJR-type algorithms," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 848–859, May 2001.
- [8] X. Chen and K. M. Chugg, "Reduced-state soft-input/output algorithms for complexity reduction in iterative and non-iterative data detection," in *Proc. IEEE Int. Conf. Communications*, New Orleans, LA, Jun. 2000, pp. 6–10.
- [9] A. G. Lillie, A. R. Nix, and J. P. McGeehan, "Performance and design of a reduced complexity iterative equalizer for precoded ISI channels," in *Proc. IEEE Vehicular Technology Conf.*, Orlando, FL, Oct. 2003, pp. 299–303.
- [10] M. Y. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 13–20, Jan. 1988.
- [11] R. E. Kamel and Y. Bar-Ness, "Reduced-complexity sequence estimation using state partitioning," *IEEE Trans. Commun.*, vol. 44, no. 9, pp. 1057–1063, Sep. 1996.
- [12] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 428–436, May 1989.
- [13] B. J. Frey and R. Kschischang, "Early detection and trellis splicing: Reduced-complexity iterative decoding," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 153–159, Feb. 1998.
- [14] V. Franz and J. B. Anderson, "Concatenated decoding with a reduced-search BCJR algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 186–195, Feb. 1998.
- [15] F. Jelinek and J. Anderson, "Instrumentable tree encoding of information sources," *IEEE Trans. Inf. Theory*, vol. 17, no. 1, pp. 118–119, Jan. 1971.
- [16] J. B. Anderson, "Limited search trellis decoding of convolutional codes," *IEEE Trans. Inf. Theory*, vol. 35, no. 9, pp. 944–955, Sep. 1989.
- [17] C. Fragouli, N. Al-Dhahir, S. N. Diggavi, and W. Turin, "Prefiltered space-time M-BCJR equalizer for frequency-selective channels," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 742–753, May 2002.
- [18] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, no. 5, pp. 779–785, May 1970.
- [19] Y. Li, B. Vucetic, and Y. Sato, "Optimum soft-output detection for channels with intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 704–713, May 1995.
- [20] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques and Software*. Dedham, MA: Artech House, 1993.
- [21] J. Luo, K. R. Pattipati, P. K. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous CDMA using probabilistic data association," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 361–363, Sep. 2001.
- [22] D. Pham, K. R. Pattipati, P. K. Willett, and J. Luo, "A generalized probabilistic data association detector for multiple antenna systems," *IEEE Commun. Lett.*, vol. 8, no. 4, pp. 205–207, Apr. 2004.
- [23] S. Liu and Z. Tian, "Near-optimum soft decision equalization for frequency selective MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 721–733, Mar. 2004.
- [24] M. S. Yee, M. Sandell, and Y. Sun, "Comparison study of single-carrier and multi-carrier modulation using iterative based receiver for MIMO system," presented at the IEEE Vehicular Technology Conf., Milan, Italy, May 2004.
- [25] C. M. Vithanage, C. Andrieu, R. J. Piechocki, and M. S. Yee, "Reduced complexity equalization of MIMO systems with a fixed-lag smoothed M-BCJR algorithm," in *Proc. 6th IEEE Workshop in Signal Processing Advances in Wireless Communications*, New York, Jun. 2005, pp. 155–159.
- [26] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 7, pp. 1293–1302, Jul. 1993.
- [27] P. Robertson, P. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *Proc. IEEE Int. Conf. Communications*, Seattle, WA, Jun. 1995, pp. 1009–1013.
- [28] M. Tuchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using *a priori* information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.
- [29] P. Del Moral, *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems With Applications*. New York: Springer-Verlag, 2004.



**Cheran M. Vithanage** (S'02) was born in Colombo, Sri Lanka, in 1977. He received the B.Sc. (Eng.) (with first class honors) and the M.Sc. degrees in electronic and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2002 and 2003, respectively.

He is currently pursuing the Ph.D. degree in statistics at the Department of Mathematics, University of Bristol, Bristol, U.K., where he has developed several reduced complexity symbol detection algorithms for digital communication systems. His current research

interests include reduced-complexity trellis-based detection and MIMO symbol detection, probabilistic graphical models, and approximate inference methods.

**Christophe Andrieu** was born in France in 1968. He received the M.S. degree from the Institut National des Télécommunications, Paris, France, in 1993, and the D.E.A. and Ph.D. degrees from the University of Paris XV, Cergy—Pontoise, France, in 1994 and 1998, respectively.

From 1998 until 2000, he was a Research Associate with the Signal Processing Group, Cambridge University, Cambridge, U.K., and a College Lecturer at Churchill College, Cambridge. Since 2001, he has been a Lecturer of statistics at the Department of Mathematics, University of Bristol, Bristol, U.K. His research interests include Bayesian estimation, model selection, Markov chain Monte Carlo methods, sequential Monte Carlo methods (particle filter), stochastic algorithms for optimization with applications to the identification of hidden Markov models, spectral analysis, speech enhancement, source separation, neural networks, communications, and nuclear science, among others.



**Robert J. Piechocki** (M'01) received his M.Sc. degree in electrical engineering from the Technical University of Wroclaw, Wroclaw, Poland, in 1997, and the Ph.D. degree in electrical engineering from the University of Bristol, Bristol, U.K., in 2002.

He is currently a Research Fellow at the Centre for Communications Research, University of Bristol. His research interests lie in the areas of statistical signal processing for communications, analog VLSI signal processing, and the optimization of wireless systems.